Discrete Mathematics 72 (1988) 141–149
North-Holland

TRIAD COUNT STATISTICS

Ove FRANK

Department of Statistics, University of Stockholm, S-10691 Stockholm, Sweden

Received 21 August 1986

The triad counts of a graph G are the numbers of various distinct induced subgraphs of order 3. If G is an undirected graph, there are 4 triad counts, and if G is a digraph, there are 16 triad counts. A multigraph is a sequence $G = (G_1, \ldots, G_r)$ of graphs and digraphs defined on a common vertex set. The concept of triad counts is generalized to multigraphs with colored vertices, edges, and arcs. It is shown how triad counts in multigraphs can be used in various kinds of statistical analyses of graph data. In particular, probability distributions are investigated of the triad counts in random multigraphs.

1. Introduction

A dyad of a graph of digraph is an induced subgraph of order 2, and a triad is an induced subgraph of order 3. For a graph or digraph G of order n there are $\binom{n}{2}$ dyads and $\binom{n}{3}$ triads that can be partitioned into different equivalence classes according to isomorphism. If G is a graph, there are at most 4 nonisomorphic triads (Fig. 1), and if G is a digraph, there are at most 16 nonisomorphic triads (Fig. 2). The frequencies of isomorphic dyads and triads in G are called the dyad and triad counts of G.

In statistical applications of random graph models the dyad and triad counts have been found useful both as summary statistics for exploratory analyses and as sufficient statistics for inference on particular random graph models. Some further comments and references on the use of dyad and triad counts in statistics are given in the next section. This shows that there is a need for extending the concept of dyad and triad counts to multigraphs $G = (G_1, \ldots, G_r)$, that is to ordered sequences of graphs and digraphs on a common vertex set. Various other extensions such as signed graphs, valued graphs and colored graphs are also discussed. Section 3 gives formulae for the numbers of nonisomorphic dyads and triads that exist for colored multigraphs. Section 4 discusses properties of the triad counts and some open extremal graph problems of interest in this connection. Section 5 gives some probabilistic properties of the triad counts for a random graph model that can be used for colored multigraphs.

2. Statistical applications

Data to be modelled by random graphs can be obtained from social contact networks, paired comparison experiments, political dominance patterns, unreli-

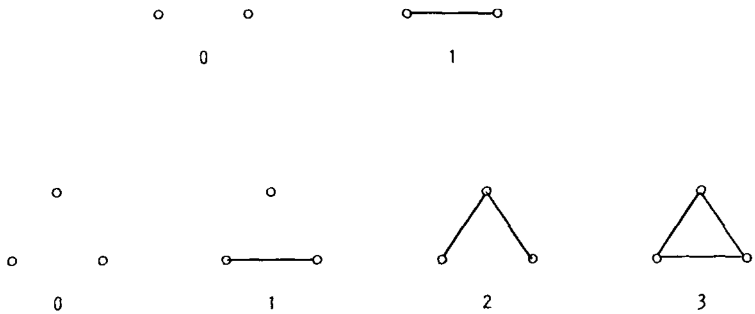

Fig. 1.

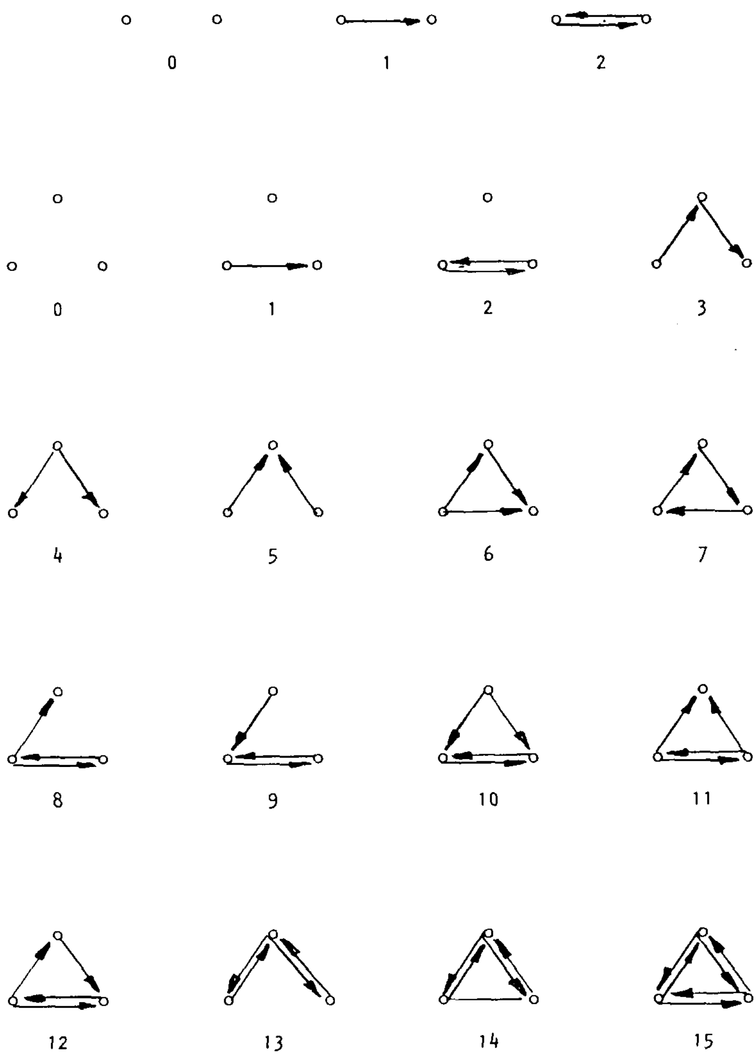

Fig. 2.

able component systems, and so forth. Often there is more than a single binary relationship to be investigated, and often there are variables defined both on the objects and on the pairs of objects involved. For instance, the vertices of a graph model can be people with vertex variables gender, age, income, etc. and edges and arcs for contacts having contact variables duration, purposes, gain, etc. Exploratory statistical analyses of graph data can be based on summary statistics like the triad counts. If different values on vertex, edge, and arc variables are represented by colors we can speak of triad counts in colored multigraphs. For instance, 4 arc colors are needed to represent the occurrences and non-occurrences of two different kinds of relationship, and 3 arc colors are needed to represent a single relationship that might be strong, weak, or absent.

Several authors have used triad counts to investigate structural properties of empirical networks and to estimate and test particular random graph models. The 16 triad counts in digraphs have been extensively studied in social science applications. See for instance Holland and Leinhardt [15, 16] and other articles in the same proceedings volumes. Transitivity indices defined in terms of triad counts were considered by Frank [6], Frank and Harary [10] and Holland and Leinhardt [13, 14, 15]. Frank [3, 4] used triad counts to study transitivity and clustering properties of graphs. Frank and Harary [8] used triad counts in signed graphs to investigate structural balance. Frank and Strauss [11] proved that the triad counts are sufficient statistics for a particular class of Markov random graph models.

We can distinguish between two different main uses of triad counts in statistical analysis. The triad counts can be used for estimating and testing particular random graph models, for instance estimating a Markov model or testing a pure randomness model versus a Markov model. A pure randomness model can here be a Bernoulli graph or a dyad independence model of the type that is described in Section 5.

Triad counts can also be used in exploring and modelling graph data, for instance by using normalized triad counts as regressors for predicting graph properties or as discriminators for graph classification. By normalized triad counts is here meant observed triad counts divided by their estimated expected values according to some random graph model. Several examples of these uses can be found in the references given above.

3. Numbers of dyads and triads

The appropriate choice and scaling of variables for multivariate graph data can be partly guided by considering the effects on the numbers of dyads and triads, that is on the number of statistics to be used in the subsequent analysis. In order to determine these numbers in general we introduce a multigraph with colored vertices, edges and arcs.

Theorem 1. *For multigraphs with a vertex colors, b edge colors and c arc colors there are* $\binom{ac+1}{2}b$ *nonisomorphic dyads and*

$$\binom{abc^2+2}{3} - a^2b^2c^2\binom{c}{2}$$

nonisomorphic triads.

Proof. Burnside's lemma (see for instance Harary [12], p. 181) can be applied to count the numbers of nonisomorphic dyads and triads. The number of nonisomor-

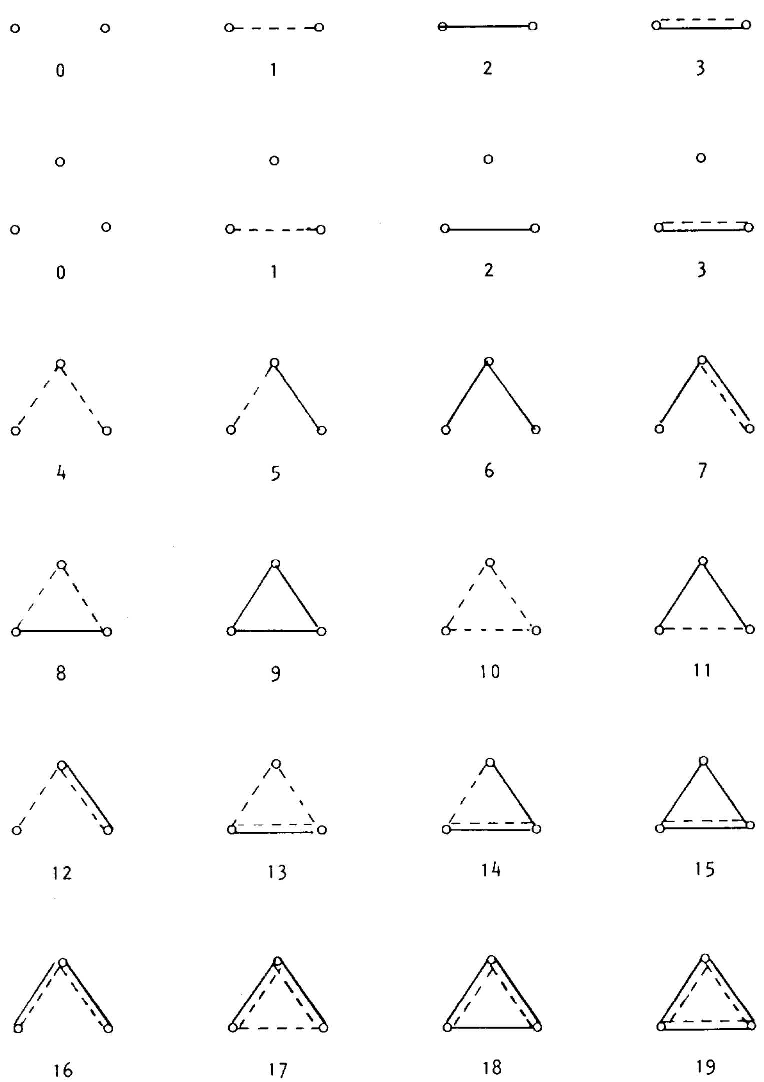

Fig. 3.

phic dyads is given by the average number of fixpoints of the vertex permutations applied to all a^2bc^2 labeled dyads that can be formed with the colors. For the identity permutation all labeled dyads are fix-points, and for the switch permutation there are abc fixpoints. Thus the number of nonisomorphic dyads is given by $\frac{1}{2}(a^2bc^2 + abc) = \binom{ac+1}{2}b$. For triads there are six vertex permutations: the identity, three switches, and two rotations. The numbers of fixpoints are found to be $a^3b^3c^6$ for the identity, $a^2b^2c^3$ for any switch, and abc^2 for any rotation. Thus the number of nonisomorphic triads is given by the average number of fixpoints

$$\frac{(a^3b^3c^6 + 3a^2b^2c^3 + 2abc^2)}{6} = \binom{abc^2+2}{3} - a^2b^2c^2\binom{c}{2}. \quad \square$$

By putting $c = 1$ and $b = 1$, respectively, we obtain the following two corollaries.

Corollary 1.1. *For graphs with a vertex colors and b edge colors there are* $\binom{a+1}{2}b$ *distinct dyads and* $\binom{ab+2}{3}$ *distinct triads.*

Corollary 1.2. *For digraphs with a vertex colors and c arc colors there are* $\binom{ac+1}{2}$ *distinct dyads and* $\binom{ac^2+2}{3} - a^2c^2\binom{c}{2}$ *distinct triads.*

By putting $a = 1$, $b = 2^g$ and $c = 2^d$ we obtain the following corollary.

Corollary 1.3. *For multigraphs consisting of g graphs and d digraphs there are* $(2^d + 1)2^{d+g-1}$ *distinct dyads and* $\binom{2^{g+2d}+1}{3} + 2^d\binom{2^{g+d}+1}{2}$ *distinct triads.*

In particular, there are 20 triads (Fig. 3) for multigraphs consisting of two graphs, and there are 104 triads for multigraphs consisting of a graph and a digraph.

4. Dyad and triad counts

Consider colored multigraphs with a vertex colors, b edge colors and c arc colors. Denote the dyad counts of a colored multigraph G by $(r_0, \ldots, r_k)$, where $k + 1$ is the number of nonisomorphic dyads, and these dyads have been labeled according to some fixed order. The triad counts of G are denoted by $(t_0, \ldots, t_l)$, where $l + 1$ is the number of nonisomorphic triads also labeled according to some fixed order. If G has order n, then

$$\sum_{i=0}^{k} r_i = \binom{n}{2}, \qquad \sum_{j=0}^{l} t_j = \binom{n}{3}.$$

The dyad counts can be obtained from the triad counts. In fact, if r_{ij} is the number of i-dyads in the j-triad, then

$$\sum_{j=0}^{l} r_{ij}t_j = (n-2)r_i, \quad i = 0, \ldots, k.$$

Various properties of G can be expressed in terms of these counts. For instance, for a single graph G take r_i and t_j as the numbers of dyads of size i for $i = 0, 1$ and triads of size j for $j = 0, 1, 2, 3$. If G has order n, size r, degrees $d_1, \ldots, d_n$ and s two-paths, then by using that

$$s = t_2 + 3t_3 = \sum_{i=1}^{n} \binom{d_i}{2}$$

we easily find that the mean and variance of the degrees can be obtained from the triad counts according to

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{2(t_1 + 2t_2 + 3t_3)}{n(n-2)},$$

$$\frac{1}{n}\sum_{i=1}^{n} (d_i - \bar{d})^2 = \bar{d}(1-\bar{d}) + \frac{2(t_2 + 3t_3)}{n}.$$

A problem of general interest to statistical applications is to find upper and lower bounds to the triad counts of colored multigraphs G with specified dyad counts. For instance we easily find that $t_3 \leq (n-2)\frac{1}{3}r$ for graphs of order n and size r. Can better bounds be found if the graph has special properties like being planar, connected, transitive, and so forth? Frank [7] considered triads in planar graphs. For instance, it holds that $t_3 \leq (n-2)^2$ for planar graphs of order n. For digraphs of odd order n, there cannot be more than $\frac{1}{4}\binom{n+1}{3}$ 3-cycles. Frank and Harary [9] considered maximum triad counts in graphs and digraphs of order n, and Bollobás [1] has a lot of material on related issues. For instance, for graphs of order n and size r

$$\max t_3 = \binom{v}{3} + \binom{r - \binom{v}{2}}{2}$$

where v is given by $\binom{v}{2} \leq r < \binom{v+1}{2}$. A lower bound to t_3 for $r \geq \frac{1}{4}n^2$ is given by $r(4r - n^2)/3n$, and a lower bound to t_3 for $\frac{1}{4}n^2 \leq r \leq \frac{1}{3}n^2$ is given by $n(4r - n^2)/9$ (Bollobás [1], Corollary 1.6 p. 297 and Corollary 1.9 p. 301). For $n \to \infty$ and $r/\binom{n}{2} \to p$ where $\frac{1}{2} \leq p \leq \frac{2}{3}$ we obtain that

$$2(2p-1)/3 \leq t_3 \Big/ \binom{n}{3} \leq p^{3/2}.$$

A limiting version of interest of the problem for colored multigraphs is to let $n \to \infty$ so that the relative dyad counts tend to specified limits $r_i/\binom{n}{2} \to p_i$ for $i = 0, \ldots, k$ and ask for bounds to the relative triad counts $t_j/\binom{n}{3}$ for $j = 0, \ldots, l$.

5. Random colored multigraphs

In random graph theory the most common models are the uniform random graph of fixed size and the Bernoulli graph on a fixed finite vertex set, say $[n] = \{1, \ldots, n\}$. A uniform random graph of size r and order n is a random graph that assigns the same probability $\binom{n}{r}^{-1}$ to each graph of size r on $[n]$. A Bernoulli graph of order n with edge probability p is a random graph that assigns probability

$$p^r(1-p)^{\binom{n}{2}-r}$$

to each graph of size r on $[n]$ for $r = 0, \ldots, \binom{n}{2}$. These models are treated extensively in the recent monographs by Bollobás [2] and Palmer [17].

A natural extension of the Bernoulli-graph to digraphs is the dyad independence model considered by Frank [6]. According to this model the $\binom{n}{2}$ random dyads are independent identically distributed with a probability distribution that is invariant under isomorphism. Thus, the random digraph distribution is specified by the probabilities assigned to the dyads of size i for $i = 0, 1, 2$.

For a colored multigraph with no vertex colors we can define a similar general dyad independence model by specifying the probabilities p_i assigned to the nonisomorphic dyads $i = 0, \ldots, k$. Here $p_i \geq 0$ and $p_0 + \cdots + p_k = 1$. Each labeled colored multigraph on $[n]$ with dyad counts $(r_0, \ldots, r_k)$ is assigned a probability

$$\prod_{i=0}^{k} p_i^{r_i}.$$

Let $(R_0, \ldots, R_k)$ be the random dyad counts and $(T_0, \ldots, T_l)$ the random triad counts of a general dyad independence model. Now $(R_1, \ldots, R_k)$ is multinomially distributed with parameters $\binom{n}{2}$ and $(p_1, \ldots, p_k)$. The distribution of the triad counts is not easy to determine. However, the expected values and the variances and covariances can be obtained by generalizing the methods used by Frank [5]. We have the following result.

Theorem 2. *For a random colored multigraph of order n the triad counts $(T_0, \ldots, T_l)$ have expected values $ET_i = \binom{n}{3}P_i$ and covariances*

$$Cov(T_i, T_j) = \binom{n}{3}P_i(\delta_{ij} - P_j) + 12\binom{n}{4}(P_{ij} - P_iP_j)$$

where P_i is the probability that a random triad is isomorphic to the i-triad, P_{ij} is the probability that two random triads with two vertices in common are isomorphic to the i-triad and the j-triad, resp., and δ_{ij} is 1 or 0 according to whether or not $i = j$.

In order to determine P_i it is convenient to use that

$$P_i = \frac{3!}{r_{0i}! \cdots r_{ki}!} p_0^{r_{0i}} \cdots p_k^{r_{ki}} w_i$$

Table 1. Triad probabilities of digraphs.
(Dyads and triads labeled according to Fig. 2.)

Triad i	Dyad counts $r_{0i}r_{1i}r_{2i}$	Labelings	Proportion w_i	Probability P_i
0	300	1	1	p_0^3
1	210	6	1	$3p_0^2p_1$
2	201	3	1	$3p_0^2p_2$
3	120	6	$\frac{1}{2}$	$3p_0p_1^2/2$
4	120	3	$\frac{1}{4}$	$3p_0p_1^2/4$
5	120	3	$\frac{1}{4}$	$3p_0p_1^2/4$
6	030	6	$\frac{3}{4}$	$3p_1^3/4$
7	030	2	$\frac{1}{4}$	$p_1^3/4$
8	111	6	$\frac{1}{2}$	$3p_0p_1p_2$
9	111	6	$\frac{1}{2}$	$3p_0p_1p_2$
10	021	3	$\frac{1}{4}$	$3p_1^2p_2/4$
11	021	3	$\frac{1}{4}$	$3p_1^2p_2/4$
12	021	6	$\frac{1}{2}$	$3p_1^2p_2/2$
13	102	3	1	$3p_0p_2^2$
14	012	6	1	$3p_1p_2^2$
15	003	1	1	p_2^3

Table 2. Triad probabilities of multigraphs consisting of two graphs. (Dyads and triads labeled according to Fig. 3.)

Triad i	Dyad counts $r_{0i}r_{1i}r_{2i}r_{3i}$	Labelings	Proportion w_i	Probability P_i
0	3000	1	1	p_0^3
1	2100	3	1	$3p_0^2p_1$
2	2010	3	1	$3p_0^2p_2$
3	2001	3	1	$3p_0^2p_3$
4	1200	3	1	$3p_0p_1^2$
5	1110	6	1	$6p_0p_1p_2$
6	1020	3	1	$3p_0p_2^2$
7	1011	6	1	$6p_0p_2p_3$
8	0210	3	1	$3p_1^2p_2$
9	0030	1	1	p_2^3
10	0300	1	1	p_1^3
11	0120	3	1	$3p_1p_2^2$
12	1101	6	1	$6p_0p_1p_3$
13	0201	3	1	$3p_1^2p_3$
14	0111	6	1	$6p_1p_2p_3$
15	0021	3	1	$3p_2^2p_3$
16	1002	3	1	$3p_0p_3^2$
17	0102	3	1	$3p_1p_3^2$
18	0012	3	1	$3p_2p_3^2$
19	0003	1	1	p_3^3

where $(r_{0i}, \ldots, r_{ki})$ are the dyad counts of the i-triad, and w_i is the proportion of triads isomorphic to the i-triad among the triads which have the same dyad counts as the i-triad. In particular, $w_i = 1$ if the i-triad is uniquely specified by its dyad counts. Otherwise there are several non-isomorphic triads which have the same dyad counts as the i-triad. For each such triad we find how many isomorphic labelings there are and then obtain w_i as the proportion of labelings for the i-triad.

Consider for example a single digraph. Dyads and triads are given in Fig. 2. The dyad counts of the triads, the numbers of isomorphic labelings, and the probabilities are given in Table 1. Table 2 gives corresponding quantities for a multigraph consisting of two graphs. We note that when $c = 1$ all nonisomorphic triads are uniquely specified by their dyad counts.

The main use of Theorem 2 in exploratory statistics is that it shows how the triad counts can be normalized to define regressors and discriminators. We normalize according to $X_i = T_i/\binom{n}{3}\hat{P}_i$ where $\hat{P}_i$ is obtained from P_i by substituting maximum likelihood estimators for the dyad probabilities, that is p_j is replaced by $\hat{p}_j = R_j/\binom{n}{2}$ for $j = 0, \ldots, k$. The probabilistic properties of $(X_1, \ldots, X_l)$ are not well known.

References

[1] B. Bollobás, Extremal graph theory (Academic Press, London, 1978).

[2] B. Bollobás, Random graphs (Academic Press, London, 1985).

[3] O. Frank, Inferences concerning cluster structure. Proceedings in Computational Statistics, edited by L.C.A. Corsten and J. Hermans, (Physica-Verlag, Wien, 1978) 259–265.

[4] O. Frank, Estimating a graph from triad counts. Statistical Computation and Simulation 9 (1979) 31–46.

[5] O. Frank, Moment properties of subgraph counts in stochastic graphs. Annals of the New York Academy of Sciences 319 (1979) 207–218.

[6] O. Frank, Transitivity in stochastic graphs and digraphs. J. Mathemat. Sociology 7 (1980) 199–213.

[7] O. Frank, On random planar graphs. The theory and applications of graphs, edited by G. Chartrand et al. (Wiley, New York, 1981) 363–372.

[8] O. Frank and F. Harary, Balance in stochastic signed graphs. Social Networks 2 (1979) 155–164.

[9] O. Frank and F. Harary Maximum triad counts in graphs and digraphs. J. Combinatorics, Information and System Sciences 5 (1980) 1–9.

[10] O. Frank and F. Harary, Cluster inference by using transitivity indices in empirical graphs. J. Amer. Statistical Assoc. 77 (1982) 835–840.

[11] O. Frank and D. Strauss, Markov graphs. J. Amer. Statistical Assoc. 81 (1986) 832–842.

[12] F. Harary, Graph theory (Addison-Wesley, Reading, 1969).

[13] P. Holland and S. Leinhardt, A Method for detecting structure in sociometric data. Amer. J. Sociology 76 (1970) 492–513.

[14] P. Holland and S. Leinhardt, Transitivity in structural models of small groups. Comparative group studies 2 (1971) 107–124.

[15] P. Holland, and S. Leinhardt, Local structure in social networks. Sociological methodology edited by D. Heise (Jossey-Bass, San Francisco, 1975) 1–45.

[16] P. Holland and S. Leinhardt, Structural sociometry. Perspectives on social network research, edited by P. Holland and S. Leinhardt (Academic Press, New York, 1979) 63–83.

[17] E. Palmer, Graphical evolution (Wiley, New York, 1985).